Sharp large deviation results for sums of independent random variables

Xiequan Fan [a,b,*], Ion Grama[c], Quansheng Liu[c,d]

[a] *Regularity Team, INRIA Saclay, Palaiseau, 91120, France*
[b] *MAS Laboratory, Ecole Centrale Paris, 92295 Châtenay-Malabry, France*
[c] *LMBA, UMR 6205, Univ. Bretagne-Sud, Vannes, 56000, France*
[d] *College of Mathematics and Computing Science, Changsha University of Science and Technology, Changsha, 410004, China*

Abstract

We show sharp bounds for probabilities of large deviations for sums of independent random variables satisfying Bernstein's condition. One such bound is very close to the tail of the standard Gaussian law in certain case; other bounds improve the inequalities of Bennett and Hoeffding by adding missing factors in the spirit of Talagrand (1995). We also complete Talagrand's inequality by giving a lower bound of the same form, leading to an equality. As a consequence, we obtain large deviation expansions similar to those of Cramér (1938), Bahadur-Rao (1960) and Sakhanenko (1991). We also show that our bound can be used to improve a recent inequality of Pinelis (2014).

Keywords: Bernstein's inequality, sharp large deviations, Cramér large deviations, expansion of Bahadur-Rao, sums of independent random variables, Bennett's inequality, Hoeffding's inequality

2000 MSC: primary 60G50; 60F10; secondary 60E15, 60F05

---

1. Introduction

Let $\xi_1, ..., \xi_n$ be a finite sequence of independent centered random variables (r.v.'s). Denote by

$$S_n = \sum_{i=1}^{n} \xi_i \quad \text{and} \quad \sigma^2 = \sum_{i=1}^{n} \mathbf{E}[\xi_i^2]. \tag{1}$$

Starting from the seminal work of Cramér [13] and Bernstein [10], the estimation of the tail probabilities $\mathbf{P}\left(S_n > x\right)$, for large $x > 0$, has attracted much attention. Various precise inequalities and asymptotic results have been established by Hoeffding [25], Nagaev [32], Saulis and Statulevicius [41], Chaganty and Sethuraman [12] and Petrov [35] under different backgrounds.

Assume that $(\xi_i)_{i=1,...,n}$ satisfies Bernstein's condition

$$|\mathbf{E}[\xi_i^k]| \leq \frac{1}{2} k! \varepsilon^{k-2} \mathbf{E}[\xi_i^2], \quad \text{for} \ \ k \geq 3 \ \ \text{and} \ \ i = 1, ..., n, \tag{2}$$

for some constant $\varepsilon > 0$. By employing the exponential Markov inequality and an upper bound for the moment generating function $\mathbf{E}[e^{\lambda \xi_i}]$, Bernstein [10] (see also Bennett [3]) has obtained the following inequalities: for all $x \geq 0$,

$$\mathbf{P}(S_n > x\sigma) \quad \leq \quad \inf_{\lambda \geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}] \tag{3}$$

---

*Corresponding author.
E-mail: fanxiequan@hotmail.com (X. Fan), ion.grama@univ-ubs.fr (I. Grama), quansheng.liu@univ-ubs.fr (Q. Liu).

$$
\begin{aligned}
&\le \quad B\left(x, \frac{\varepsilon}{\sigma}\right) := \exp\left\{-\frac{\widehat{x}^2}{2}\right\} &(4)\\
&\le \quad \exp\left\{-\frac{x^2}{2(1+x\varepsilon/\sigma)}\right\}, &(5)
\end{aligned}
$$

where

$$
\widehat{x} = \frac{2x}{1+\sqrt{1+2x\varepsilon/\sigma}}; \tag{6}
$$

see also van de Geer and Lederer [47] with a new method based on Bernstein-Orlicz norm and Rio [40]. Some extensions of the inequalities of Bernstein and Bennett can be found in van de Geer [46] and de la Peña [14] for martingales; see also Rio [38, 39] and Bousquet [11] for the empirical processes with r.v.'s bounded from above.

Since $\lim_{\varepsilon/\sigma\to 0}\mathbf{P}(S_n > x\sigma) = 1-\Phi(x)$ and $\lim_{\varepsilon/\sigma\to 0} B\left(x, \frac{\varepsilon}{\sigma}\right) = e^{-x^2/2}$, where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$ is the standard normal distribution function, the central limit theorem (CLT) suggests that Bennett's inequality (4) can be substantially refined by adding the factor

$$
M(x) = \Big(1-\Phi(x)\Big)\exp\left\{\frac{x^2}{2}\right\},
$$

where $\sqrt{2\pi}M(x)$ is known as Mill's ratio. It is known that $M(x)$ is of order $1/x$ as $x\to\infty$.

To recover a factor of order $1/x$ as $x\to\infty$ a lot of effort has been made. Certain factors of order $1/x$ have been recovered by using the following inequality: for some $\alpha > 1$,

$$
\mathbf{P}(S_n \ge x\sigma) \quad \le \quad \inf_{t<x\sigma}\mathbf{E}\left[\frac{((S_n-t)^+)^\alpha}{((x\sigma-t)^+)^\alpha}\right],
$$

where $x^+ = \max\{x, 0\}$; see Eaton [17], Bentkus [4], Pinelis [36] and Bentkus et al. [7]. Some bounds on tail probabilities of type

$$
\mathbf{P}(S_n \ge x\sigma) \;\le\; C\Big(1-\Phi(x)\Big), \tag{7}
$$

where $C > 1$ is an absolute constant, are obtained for sums of weighted Rademacher r.v.'s; see Bentkus [4]. In particular, Bentkus and Dzindzalieta [6] proved that

$$
C = \frac{1}{4(1-\Phi(\sqrt{2}))} \approx 3.178
$$

is sharp in (7).

When the summands ξ_i are bounded from above, results of such type have been obtained by Talagrand [45], Bentkus [5] and Pinelis [37]. Using the conjugate measure technique, Talagrand (cf. Theorems 1.1 and 3.3 of [45]) proved that if the r.v.'s satisfy $\xi_i \le 1$ and $|\xi_i| \le b$ for a constant $b > 0$ and all $i = 1, ..., n$, then there exists an universal constant K such that, for all $0 \le x \le \frac{\sigma}{Kb}$,

$$
\begin{aligned}
\mathbf{P}(S_n > x\sigma) \quad &\le \quad \inf_{\lambda\ge 0}\mathbf{E}[e^{\lambda(S_n-x\sigma)}]\left(M(x)+K\frac{b}{\sigma}\right) &(8)\\
&\le \quad H_n(x,\sigma)\left(M(x)+K\frac{b}{\sigma}\right), &(9)
\end{aligned}
$$

where

$$
H_n(x,\sigma) = \left\{\left(\frac{\sigma}{x+\sigma}\right)^{x\sigma+\sigma^2}\left(\frac{n}{n-x\sigma}\right)^{n-x\sigma}\right\}^{\frac{n}{n+\sigma^2}}.
$$

Since $M(x) = O\left(\frac{1}{x}\right), x \to \infty$, equality (9) improves on Hoeffding's bound $H_n(x, \sigma)$ (cf. (2.8) of [25]) by adding a missing factor

$$F_1\Big(x, \frac{b}{\sigma}\Big) = M(x) + K\frac{b}{\sigma}$$

of order $\frac{1}{x}$ for the range $0 \leq x \leq \frac{\sigma}{Kb}$. Other improvements on Hoeffding's bound can be found in Bentkus [5] and Pinelis [36]. Bentkus's inequality [5] is much better than (9) in the sense that it recovers a factor of order $\frac{1}{x}$ for all $x \geq 0$ instead of the range $0 \leq x \leq \frac{\sigma}{Kb}$, and do not assume that ξ_i's have moments of order larger than 2; see also Pinelis [37] for a similar improvement on Bennett-Hoeffding's bound.

The scope of this paper is to give several improvements on Bernstein's inequalities (3), (5) and Bennett's inequality (4) for sums of non-bounded r.v.'s instead of sums of bounded (from above) r.v.'s, which are considered in Talagrand [45], Bentkus [5] and Pinelis [36]. Moreover, some tight lower bounds are also given, which were not considered by Talagrand [45], Bentkus [5] and Pinelis [36]. In particular, we improve Talagrand's inequality to an *equality*, which will imply simple large deviation expansions. We also show that our bound can be used to improve a recent upper bound on tail probabilities due to Pinelis [36].

Our approach is based on the conjugate distribution technique due to Cramér, which becomes a standard for obtaining sharp large deviation expansions. We refine the technique inspired by Talagrand [45] and Grama and Haeusler [23] (see also [19, 20]), and derive sharp bounds for the cumulant function to obtain precise upper bounds on tail probabilities under Bernstein's condition.

As to the potential applications of our results in statistics, we refer to Fu, Li and Zhao [21] for large sample estimation and Joutard [28, 29] for nonparametric estimation. In these papers, many interesting Bahadur-Rao type large deviation expansions have been established. Our result leads to simple large deviation expansions which are similar (but simpler) to those of Cramér (1938), Bahadur-Rao (1960) and Sakhanenko (1991). For other important applications, we refer to Shao [44] and Jing, Shao and Wang [26], where the authors have established the Cramér type self-normalized large deviations for normalized $x = o(n^{1/6})$; see also Jing, Liang and Zhou [27]. From the proofs of theorems in [44, 26, 27], we find that the self-normalized large deviations are closely related to the large deviations for sums of bounded from above r.v.'s (cf. [18]). Our results may help extend the Cramér type self-normalized large deviations to a larger range.

The paper is organized as follows. In Section 2, we present our main results. In Section 3, some comparisons are given. In Section 4, we state some auxiliary results to be used in the proofs of theorems. Sections 5 - 7 are devoted to the proofs of main results.

2. Main results

All over the paper $\xi_1, ..., \xi_n$ is a finite sequence of independent real-valued r.v.'s with $\mathbf{E}[\xi_i] = 0$ and satisfying Bernstein's condition (2), S_n and σ^2 are defined by (1). We use the notations $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$ and $a^+ = a \vee 0$. Throughout this paper, C stands for an absolute constant with possibly different values in different places.

Our first result is the following large deviation inequality valid for all $x \geq 0$.

Theorem 2.1. *For any $\delta \in (0, 1]$ and $x \geq 0$,*

$$\mathbf{P}(S_n > x\sigma) \leq \Big(1 - \Phi\left(\widetilde{x}\right)\Big)\left[1 + C_\delta\left(1 + \widetilde{x}\right)\frac{\varepsilon}{\sigma}\right], \tag{10}$$

where

$$\widetilde{x} = \frac{2x}{1 + \sqrt{1 + 2(1+\delta)x\varepsilon/\sigma}} \tag{11}$$

and C_δ is a constant only depending on δ. In particular, if $0 \leq x = o(\sigma/\varepsilon)$, $\varepsilon/\sigma \to 0$, then

$$\mathbf{P}(S_n > x\sigma) \leq \Big(1 - \Phi\left(\widetilde{x}\right)\Big)\Big[1 + o(1)\Big].$$

The interesting feature of the bound (10) is that it decays exponentially to 0 and also recovers closely the shape of the standard normal tail $1-\Phi(x)$ when $r=\frac{\varepsilon}{\sigma}$ becomes small, which is not the case of Bennett's bound $B(x,\frac{\varepsilon}{\sigma})$ and Berry-Essen's bound

$$\mathbf{P}(S_n > x\sigma) \leq 1-\Phi\left(x\right)+C\frac{\varepsilon}{\sigma}.$$

Our result can be compared with Cramér's large deviation result in the i.i.d. case (cf. (34)). With respect to Cramér's result, the advantage of (10) is that it is valid for all $x \geq 0$.

Notice that Theorem 2.1 improves Bennett's bound only for moderate x. A further significant improvement of Bennett's inequality (4) for all $x \geq 0$ is given by the following theorem: We replace Bennett's bound $B\left(x,\frac{\varepsilon}{\sigma}\right)$ by the following smaller one:

$$B_n\left(x,\frac{\varepsilon}{\sigma}\right)=B\left(x,\frac{\varepsilon}{\sigma}\right)\exp\left\{-n\psi\left(\frac{\widehat{x}^2}{2n\sqrt{1+2x\varepsilon/\sigma}}\right)\right\}, \tag{12}$$

where $\psi(t)=t-\log(1+t)$ is a nonnegative convex function in $t \geq 0$.

Theorem 2.2. *For all* $x \geq 0$,

$$\begin{aligned}\mathbf{P}(S_n > x\sigma) \quad &\leq \quad B_n\left(x,\frac{\varepsilon}{\sigma}\right)\,F_2\left(x,\frac{\varepsilon}{\sigma}\right) &(13)\\ &\leq \quad B_n\left(x,\frac{\varepsilon}{\sigma}\right), &(14)\end{aligned}$$

where

$$F_2\left(x,\frac{\varepsilon}{\sigma}\right)=\Big(M(x)+27.99R\left(x\varepsilon/\sigma\right)\frac{\varepsilon}{\sigma}\Big)\wedge 1 \tag{15}$$

and

$$R(t)=\begin{cases}\frac{(1-t+6t^2)^3}{(1-3t)^{3/2}(1-t)^7}, & \text{if } 0\leq t<\frac{1}{3},\\ \infty, & \text{if } t\geq \frac{1}{3},\end{cases} \tag{16}$$

is an increasing function. Moreover, for all $0 \leq x \leq \alpha\frac{\sigma}{\varepsilon}$ *with* $0 \leq \alpha < \frac{1}{3}$*, it holds* $R(x\varepsilon/\sigma) \leq R(\alpha)$*. If* $\alpha = 0.1$*, then* $27.99R(\alpha) \leq 88.41$*.*

To highlight the improvement of Theorem 2.2 over Bennett's bound, we note that $B_n(x,\frac{\varepsilon}{\sigma}) \leq B(x,\frac{\varepsilon}{\sigma})$ and, in the i.i.d. case (or, more generally when $\frac{\varepsilon}{\sigma}=\frac{c_0}{\sqrt{n}}$, for some constant $c_0>0$),

$$B_n\left(\sqrt{n}x,\frac{\varepsilon}{\sigma}\right)=B\left(\sqrt{n}x,\frac{\varepsilon}{\sigma}\right)\exp\left\{-c_x\, n\right\}, \tag{17}$$

where $c_x > 0$, $x > 0$, does not depend on n. Thus Bennett's bound is strengthened by adding a factor $\exp\{-c_x\, n\}, n\to\infty$, which is similar to Hoeffding's improvement on Bennett's bound for sums of bounded r.v.'s [25]. The second improvement in the right-hand side of (13) comes from the missing factor $F_2(x,\frac{\varepsilon}{\sigma})$, which is of order $M(x)[1+o(1)]$ for moderate values of x satisfying $0 \leq x = o(\frac{\sigma}{\varepsilon})$, $\frac{\varepsilon}{\sigma}\to 0$. This improvement is similar to Talagrand's refinement on Hoeffding's upper bound $H_n(x,\sigma)$ by the factor $F_1(x,b/\sigma)$; see (9). The numerical values of the missing factor $F_2(x,\frac{\varepsilon}{\sigma})$ are displayed in Figure 1.

Our numerical results confirm that the bound $B_n(x,\frac{\varepsilon}{\sigma})F_2(x,\frac{\varepsilon}{\sigma})$ in (13) is better than Bennett's bound $B(x,\frac{\varepsilon}{\sigma})$ for all $x \geq 0$. For the convenience of the reader, we display the ratios of $B_n(x,r)F_2(x,r)$ to $B(x,r)$ in Figure 2 for various $r=\frac{1}{\sqrt{n}}$.

The following corollary improves inequality (10) of Theorem 2.1 in the range $0 \leq x \leq \alpha\frac{\sigma}{\varepsilon}$ with $0 \leq \alpha < \frac{1}{3}$. It corresponds to taking $\delta = 0$ in the definition (11) of $\widetilde{x}$.

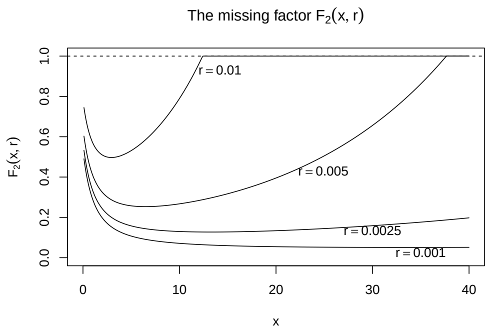

Figure 1: The missing factor $F_2(x, r)$ is displayed as a function of x for various values of $r = \frac{\varepsilon}{\sigma}$.

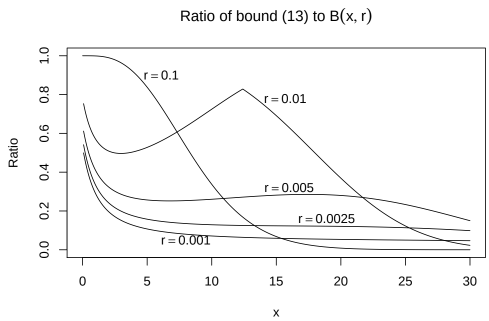

Figure 2: Ratio of $B_n\,(x, r)\,F_2\,(x, r)$ to $B(x, r)$ as a function of x for various values of $r = \frac{\varepsilon}{\sigma} = \frac{1}{\sqrt{n}}$.

Corollary 2.1. *For all* $0 \leq x \leq \alpha \frac{\sigma}{\varepsilon}$ *with* $0 \leq \alpha < \frac{1}{3}$,

$$\mathbf{P}(S_n > x\sigma) \quad \leq \quad \Big(1 - \Phi\left(\widehat{x}\right)\Big) \Big[1 + 70.17 R(\alpha) \left(1 + \widehat{x}\right) \frac{\varepsilon}{\sigma}\Big], \tag{18}$$

where $\widehat{x}$ *is defined in (6) and* $R(t)$ *by (16). In particular, for all* $0 \leq x = o(\frac{\sigma}{\varepsilon})$, $\frac{\varepsilon}{\sigma} \to 0$,

$$\begin{aligned} \mathbf{P}(S_n > x\sigma) \quad &\leq \quad \Big(1 - \Phi\left(\widehat{x}\right)\Big) \Big[1 + o(1)\Big] \\ &= \quad B\left(x, \frac{\varepsilon}{\sigma}\right) M(\widehat{x}) \Big[1 + o(1)\Big]. \end{aligned} \tag{19}$$

The advantage of Corollary 2.1 is that in the normal distribution function $\Phi(x)$ we have the expression $\widehat{x}$ instead of the smaller term $\widetilde{x}$ figuring in Theorem 2.1, which represents a significant improvement.

Notice that inequality (19) improves Bennett's bound $B\left(x, \frac{\varepsilon}{\sigma}\right)$ by the missing factor $M(\widehat{x})[1 + o(1)]$ for all $0 \leq x = o\left(\frac{\sigma}{\varepsilon}\right)$.

For the lower bound of tail probabilities $\mathbf{P}(S_n > x\sigma)$, we have the following result, which is a complement of Corollary 2.1.

Theorem 2.3. *For all* $0 \leq x \leq \alpha \frac{\sigma}{\varepsilon}$ *with* $0 \leq \alpha \leq \frac{1}{9.6}$,

$$\mathbf{P}(S_n > x\sigma) \quad \geq \quad \Big(1 - \Phi\left(\check{x}\right)\Big) \Big[1 - c_\alpha \left(1 + \check{x}\right) \frac{\varepsilon}{\sigma}\Big],$$

where $\check{x} = \frac{\lambda\sigma}{(1-\lambda\varepsilon)^3}$ *with* $\lambda = \frac{2x/\sigma}{1+\sqrt{1-9.6x\varepsilon/\sigma}}$, *and* $c_\alpha = 67.38 R\left(\frac{2\alpha}{1+\sqrt{1-9.6\alpha}}\right)$ *is a bounded function. Moreover, for all* $0 \leq x = o(\frac{\sigma}{\varepsilon})$, $\frac{\varepsilon}{\sigma} \to 0$,

$$\mathbf{P}(S_n > x\sigma) \geq \Big(1 - \Phi\left(\check{x}\right)\Big) \Big[1 - o(1)\Big].$$

Combining Corollary 2.1 and Theorem 2.3, we obtain, for all $0 \leq x \leq 0.1 \frac{\sigma}{\varepsilon}$,

$$\mathbf{P}(S_n > x\sigma) = \left(1 - \Phi\Big(x(1 + \theta_1 c_1 x \frac{\varepsilon}{\sigma})\Big)\right) \left[1 + \theta_2 c_2 (1 + x) \frac{\varepsilon}{\sigma}\right], \tag{20}$$

where $c_1, c_2 > 0$ are absolute constants and $|\theta_1|, |\theta_2| \leq 1$. This result can be found in Sakhanenko [43] but in a more narrow zone.

Some earlier lower bounds on tail probabilities, based on Cramér large deviations, can be found in Arkhangelskii [1], Nagaev [33] and Rozovky [30]. In particular, Nagaev established the following lower bound

$$\mathbf{P}(S_n > x\sigma) \quad \geq \quad \Big(1 - \Phi(x)\Big) e^{-c_1 x^3 \frac{\varepsilon}{\sigma}} \left(1 - c_2 (1 + x) \frac{\varepsilon}{\sigma}\right) \tag{21}$$

for some explicit constants c_1, c_2 and all $0 \leq x \leq \frac{1}{25} \frac{\sigma}{\varepsilon}$. For more general results, we refer to Theorem 3.1 of Saulis and Statulevicius [41].

In the following theorem, we obtain a one term sharp large deviation expansion similar to Cramér [13], Bahadur-Rao [2], Saulis and Statulevicius [41] and Sakhanenko [43].

Theorem 2.4. *For all* $0 \leq x < \frac{1}{12} \frac{\sigma}{\varepsilon}$,

$$\mathbf{P}(S_n > x\sigma) = \inf_{\lambda \geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}] \, F_3\left(x, \frac{\varepsilon}{\sigma}\right), \tag{22}$$

where

$$F_3\left(x, \frac{\varepsilon}{\sigma}\right) = M(x) + 27.99 \, \theta R\left(4x\varepsilon/\sigma\right) \frac{\varepsilon}{\sigma}, \tag{23}$$

$|\theta| \leq 1$ *and* $R(t)$ *is defined by (16). Moreover,* $\inf_{\lambda\geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}] \leq B(x, \frac{\varepsilon}{\sigma})$*. In particular, in the i.i.d. case, we have the following non-uniform Berry-Esseen type bound: for all* $0 \leq x = o(\sqrt{n})$*,*

$$\Big|\mathbf{P}(S_n > x\sigma) - M(x) \inf_{\lambda\geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}]\Big| \leq \frac{C}{\sqrt{n}} B(x, \frac{\varepsilon}{\sigma}). \tag{24}$$

Theorem 2.4 holds also for ξ_i's bounded from above. In this case the term $27.99\,\theta R\,(4x\varepsilon/\sigma)$ can be significantly refined; see [18]. In particular, if $|\xi_i| \leq \varepsilon$, then $27.99\,\theta R\,(4x\varepsilon/\sigma)$ can be improved to 3.08. However, under the stated condition of Theorem 2.4, the term $27.99\,\theta R\,(4x\varepsilon/\sigma)$ cannot be improved significantly.

When Bernstein's condition fails, we refer to Theorem 3.1 of Saulis and Statulevicius [41], where explicit and asymptotic expansions have been established via the Cramér series (cf. Petrov [34] for details). When the Bernstein condition holds, their result reduces to the result of Cramér [13]. However, they gave an explicit information on the term corresponding to our term $27.99\,\theta R\,(4x\varepsilon/\sigma)$.

Equality (22) shows that $\inf_{\lambda\geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}]$ is the best possible exponentially decreasing rate on tail probabilities. It reveals the missing factor F_3 in Bernstein's bound (3) (and thus in many other classical bounds such as Hoeffding, Bennett and Bernstein). Since $\theta \geq -1$, equality (22) completes Talagrand's upper bound (8) by giving a sharp lower bound. If ξ_i are bounded from above $\xi_i \leq 1$, it holds that $\inf_{\lambda\geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}] \leq H_n(x, \sigma)$ (cf. [25]). Therefore (22) implies Talagrand's inequality (9).

A precise large deviation expansion, as sharp as (22), can be found in Sakhanenko [43] (see also Györfi, Harremöes and Tusnády [24]). In his paper, Sakhanenko proved an equality similar to (22) in a more narrow range $0 \leq x \leq \frac{1}{200}\frac{\sigma}{\varepsilon}$,

$$\Big|\mathbf{P}(S_n > x\sigma) - \Big(1 - \Phi(t_x)\Big)\Big| \;\leq\; C\frac{\varepsilon}{\sigma} e^{-t_x^2/2}, \tag{25}$$

where

$$t_x = \sqrt{-2\ln\Big(\inf_{\lambda\geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}]\Big)}$$

is a value depending on the distribution of S_n and satisfying $|t_x - x| = O(x^2\frac{\varepsilon}{\sigma})$, $\frac{\varepsilon}{\sigma} \to 0$, for moderate x's. It is worth noting that from Sakhanenko's result, we find that the inequalities (24) and (27) hold also if $M(x)$ is replaced by $M(t_x)$.

Using the two sided bound

$$\frac{1}{\sqrt{2\pi}(1+t)} \leq M(t) \leq \frac{1}{\sqrt{\pi}(1+t)}, \qquad t \geq 0, \tag{26}$$

and

$$M(t) \sim \frac{1}{\sqrt{2\pi}(1+t)}, \qquad t \to \infty$$

(see p. 17 in Itō and MacKean [22] or Talagrand [45]), equality (22) implies that the relative errors between $\mathbf{P}(S_n > x\sigma)$ and $M(x)\inf_{\lambda\geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}]$ converges to 0 uniformly in the range $0 \leq x = o\left(\frac{\sigma}{\varepsilon}\right)$ as $\frac{\varepsilon}{\sigma} \to 0$, i.e.

$$\mathbf{P}(S_n > x\sigma) \;=\; M(x) \inf_{\lambda\geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}]\Big(1 + o(1)\Big). \tag{27}$$

Expansion (27) extends the following Cramér large deviation expansion: for $0 \leq x = o\left(\sqrt[3]{\frac{\sigma}{\varepsilon}}\right)$ as $\frac{\sigma}{\varepsilon} \to \infty$,

$$\mathbf{P}(S_n > x\sigma) \;=\; \Big(1 - \Phi\,(x)\Big)\Big[1 + o(1)\Big]. \tag{28}$$

To have an idea of the precision of expansion (27), we plot the ratio

$$\mathrm{Ratio}(x, n) = \frac{\mathbf{P}(S_n \geq x\sqrt{n})}{M(x)\inf_{\lambda\geq 0} \mathbf{E}[e^{\lambda(S_n - x\sqrt{n})}]}$$

in Figure 3 for the case of sums of Rademacher r.v.'s $\mathbf{P}(\xi_i = -1) = \mathbf{P}(\xi_i = 1) = \frac{1}{2}$. From these plots we see that the error in (27) becomes smaller as n increases.

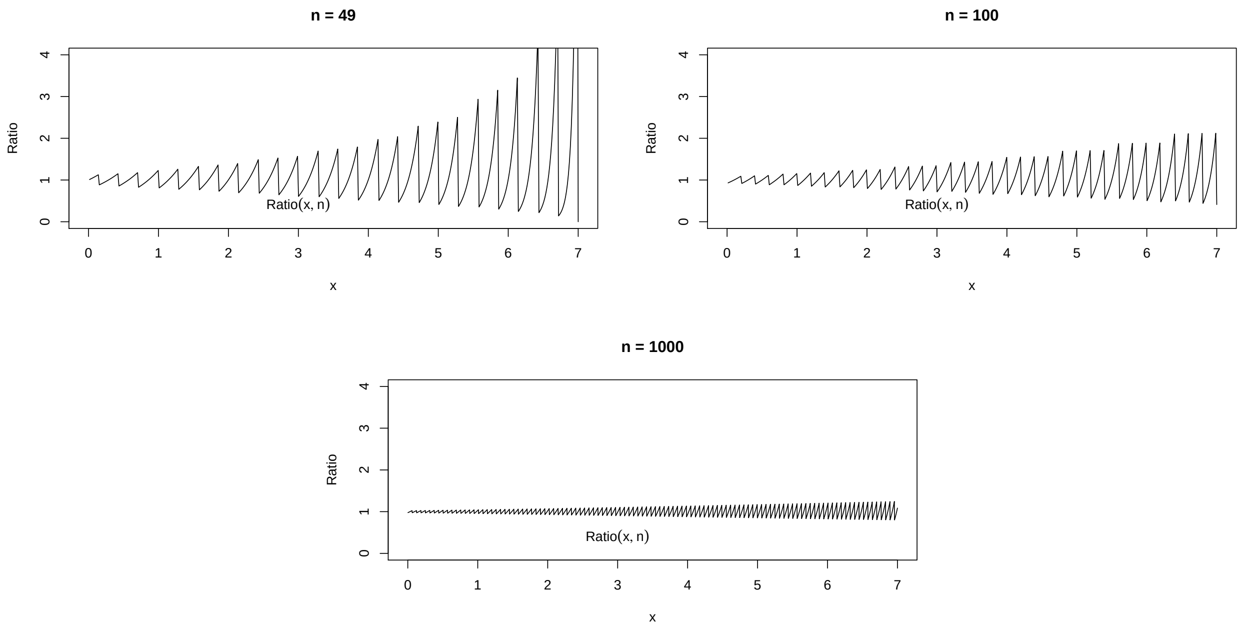

Figure 3: The ratio $\mathrm{Ratio}(x,n) = \frac{\mathbf{P}(S_n \geq x\sqrt{n})}{M(x)\inf_{\lambda\geq 0}\mathbf{E}[e^{\lambda(S_n - x\sqrt{n})}]}$ is displayed as a function of x for various n for sums of Rademacher r.v.'s.

3. Some comparisons

3.1. Comparison with a recent inequality of Pinelis

In this subsection, we show that Theorem 2.4 can be used to improve a recent upper bound on tail probabilities due to Pinelis [37]. For simplicity of notations, we assume that $\xi_i \leq 1$ and only consider the i.i.d. case. For other cases, the argument is similar. Let us recall the notations of Pinelis. Denote by Γ_{a^2} the normal r.v. with mean 0 and variance $a^2 > 0$, and Π_θ the Poisson r.v. with parameter $\theta > 0$. Let also

$$\widetilde{\Pi}_\theta \sim \Pi_\theta - \theta.$$

Denote by

$$\delta = \frac{\sum_{i=1}^n \mathbf{E}[(\xi_i^+)^3]}{\sigma^2}. \tag{29}$$

Then it is obvious that $\delta \in (0,1)$. Pinelis (cf. Corollary 2.2 of [37]) proved that: for all $y \geq 0$,

$$\mathbf{P}(S_n > y) \ \leq \ \frac{2e^3}{9}\mathbf{P}^{LC}(\Gamma_{(1-\delta)\sigma^2} + \widetilde{\Pi}_{\delta\sigma^2} > y), \tag{30}$$

where, for any r.v. ζ, the function $\mathbf{P}^{LC}(\zeta > y)$ denotes the least log-concave majorant of the tail function $\mathbf{P}(\zeta > y)$. So that $\mathbf{P}^{LC}(\zeta > y) \geq \mathbf{P}(\zeta > y)$. By the remark of Pinelis, inequality (30) refines the Bennet-Hoeffding inequality by adding a factor of order $\frac{1}{x}$ in certain range. By Theorem 2.4 and some simple calculations, we find that, for all $0 \leq y = o(n)$,

$$\mathbf{P}^{LC}(\Gamma_{(1-\delta)\sigma^2} + \widetilde{\Pi}_{\delta\sigma^2} > y)$$

$$
\begin{aligned}
&\geq \mathbf{P}(\Gamma_{(1-\delta)\sigma^2} + \widetilde{\Pi}_{\delta\sigma^2} > y) \\
&\geq \inf_{\lambda \geq 0} \mathbf{E}[e^{\lambda(\Gamma_{(1-\delta)\sigma^2} + \widetilde{\Pi}_{\delta\sigma^2} - y)}]\Big(M(y/\sigma) - \frac{C}{\sqrt{n}}\Big) \\
&= \inf_{\lambda \geq 0} \mathbf{E}[e^{-\lambda y + f(\lambda,\delta,\sigma)}]\Big(M(y/\sigma) - \frac{C}{\sqrt{n}}\Big), \qquad (31)
\end{aligned}
$$

where

$$
f(\lambda, \delta, \sigma) = \frac{\lambda^2}{2}(1-\delta)\sigma^2 + (e^{\lambda} - 1 - \lambda)\delta\sigma^2 .
$$

By the inequality

$$
e^x - 1 - x \leq \frac{x^2}{2} + \sum_{k=3}^{\infty} (x^+)^k
$$

(cf. proof of Corollary 3 in Rio [40]) and the fact that $\log(1+t)$ is concave in $t \geq 0$, it follows that, for any $\lambda > 0$,

$$
\begin{aligned}
\log \mathbf{E}[e^{\lambda S_n}] &= \sum_{i=1}^n \log \mathbf{E}[e^{\lambda \xi_i}] \leq \sum_{i=1}^n \log\Big(1 + \frac{\lambda^2}{2}\mathbf{E}[\xi_i^2] + \sum_{k=3}^{\infty} \frac{\lambda^k}{k!}\mathbf{E}[(\xi_i^+)^k]\Big) \\
&\leq \sum_{i=1}^n \log\Big(1 + \frac{\lambda^2}{2}\mathbf{E}[\xi_i^2] + \sum_{k=3}^{\infty} \frac{\lambda^k}{k!}\mathbf{E}[(\xi_i^+)^3]\Big) \\
&\leq n \log\left(1 + \frac{1}{n}\sum_{i=1}^n \Big(\frac{\lambda^2}{2}\mathbf{E}[\xi_i^2] + \sum_{k=3}^{\infty} \frac{\lambda^k}{k!}\mathbf{E}[(\xi_i^+)^3]\Big)\right) \\
&\leq n \log\left(1 + \frac{1}{n}\Big(\frac{\lambda^2}{2}\sigma^2 + \sum_{k=3}^{\infty} \frac{\lambda^k}{k!}\delta\sigma^2\Big)\right) \\
&= n \log\left(1 + \frac{1}{n} f(\lambda, \delta, \sigma)\right).
\end{aligned}
$$

By the last line, Theorem 2.4 implies that, for all $0 \leq y = o(n)$,

$$
\begin{aligned}
\mathbf{P}(S_n > y) &\leq \inf_{\lambda \geq 0} \mathbf{E}[e^{-\lambda y + n\log(1+\frac{1}{n}f(\lambda,\delta,\sigma))}]\Big(M(y/\sigma) + \frac{C}{\sqrt{n}}\Big) \\
&\leq \Big(1 + o(1)\Big) \inf_{\lambda \geq 0} \mathbf{E}[e^{-\lambda y + n\log(1+\frac{1}{n}f(\lambda,\delta,\sigma))}]\Big(M(y/\sigma) - \frac{C}{\sqrt{n}}\Big). \qquad (32)
\end{aligned}
$$

Note that $n\log\big(1 + \frac{1}{n}f(\lambda,\delta,\sigma)\big) \leq f(\lambda,\delta,\sigma)$. By the inequalities (30), (31) and (32), we find that (32) not only refines Pinelis' constant $\frac{2e^3}{9}$ (≈ 4.463) to $1 + o(1)$ for large n, but also gives an exponential bound sharper than that of Pinelis.

3.2. Comparison with the expansions of Cramér and Bahadur-Rao

Notice that the expression $\inf_{\lambda \geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}]$ can be rewritten in the form $\exp\{-n\Lambda_n^*(\frac{x\sigma}{n})\}$, where $\Lambda_n^*(x) = \sup_{\lambda \geq 0}\{\lambda x - \frac{1}{n}\log \mathbf{E}[e^{\lambda S_n}]\}$ is the Fenchel-Legendre transform of the normalized cumulant function of S_n. In the i.i.d. case, the function $\Lambda^*(x) = \Lambda_n^*(x)$ is known as the good rate function in large deviation principle (LDP) theory (see Deuschel and Stroock [16] or Dembo and Zeitouni [15]).

Now we clarify the relation among our large deviation expansion (22), Cramér large deviations [13] and the Bahadur-Rao theorem [2] in the i.i.d. case. Without loss of generality, we take $\sigma_1^2 = 1$, where σ_1^2 is the variance of ξ_1. First, our bound (22) implies that: for all $0 \leq x = o(\sqrt{n})$,

$$
\mathbf{P}\left(S_n > x\sigma\right) = e^{-n\,\Lambda^*(x/\sqrt{n})} M(x)\left[1 + O\left(\frac{1+x}{\sqrt{n}}\right)\right], \quad n \to \infty. \qquad (33)
$$

Cramér [13] (see also Theorem 3.1 of Saulis and Statulevicius [41] for more general results) proved that, for all $0 \leq x = o(\sqrt{n})$,

$$\frac{\mathbf{P}(S_n > x\sigma)}{1-\Phi(x)} = \exp\left\{ \frac{x^3}{\sqrt{n}}\lambda\left(\frac{x}{\sqrt{n}}\right)\right\}\left[1+O\left(\frac{1+x}{\sqrt{n}}\right)\right], \quad n\to\infty, \tag{34}$$

where $\lambda(\cdot)$ is the Cramér series. So the good rate function and the Cramér series have the relation $n\,\Lambda^*(\frac{x}{\sqrt{n}}) = \frac{x^2}{2} - \frac{x^3}{\sqrt{n}}\lambda\left(\frac{x}{\sqrt{n}}\right)$. Second, consider the large deviation probabilities $\mathbf{P}\left(\frac{S_n}{n} > y\right)$. Since $\frac{S_n}{n} \to 0, a.s.$, as $n\to\infty$, we only place emphasis on the case where y is small positive constant. Bahadur-Rao proved that, for given positive constant y,

$$\mathbf{P}\left(\frac{S_n}{n} > y\right) = \frac{e^{-n\,\Lambda^*(y)}}{\sigma_{1y}t_y\sqrt{2\pi n}}\left[1+O(\frac{c_y}{n})\right], \qquad n\to\infty, \tag{35}$$

where c_y, σ_{1y} and t_y depend on y and the distribution of ξ_1; see also Bercu [8, 9], Rozovky [31] and Györfi, Harremöes and Tusnády [24] for more general results. Our bound (22) implies that, for $y \geq 0$ small enough,

$$\mathbf{P}\left(\frac{S_n}{n} > y\right) = e^{-n\,\Lambda^*(y)}M(y\sqrt{n})\Big[1+O(y+\frac{1}{\sqrt{n}})\Big]. \tag{36}$$

In particular, when $0 < y = y(n) \to 0$ and $y\sqrt{n}\to\infty$ as $n\to\infty$, we have

$$\mathbf{P}\left(\frac{S_n}{n} > y\right) = \frac{e^{-n\,\Lambda^*(y)}}{y\sqrt{2\pi n}}\Big[1+o(1)\Big], \qquad n\to\infty. \tag{37}$$

Expansion (36) or (37) is less precise than (35). However, the advantage of the expansions (36) and (37) over the Bahadur-Rao expansion (35) is that the expansions (36) or (37) are uniform in y (where y may be dependent of n), in addition to the simpler expressions (without the factors t_y and σ_y).

4. Auxiliary results

We consider the positive r.v.

$$Z_n(\lambda) = \prod_{i=1}^{n}\frac{e^{\lambda\xi_i}}{\mathbf{E}[e^{\lambda\xi_i}]}, \qquad |\lambda| < \varepsilon^{-1},$$

(the Esscher transformation) so that $\mathbf{E}[Z_n(\lambda)] = 1$. We introduce the *conjugate probability measure* $\mathbf{P}_\lambda$ defined by

$$d\mathbf{P}_\lambda = Z_n(\lambda)d\mathbf{P}. \tag{38}$$

Denote by $\mathbf{E}_\lambda$ the expectation with respect to $\mathbf{P}_\lambda$. Setting

$$b_i(\lambda) = \mathbf{E}_\lambda[\xi_i] = \frac{\mathbf{E}[\xi_i e^{\lambda\xi_i}]}{\mathbf{E}[e^{\lambda\xi_i}]}, \quad i = 1, ..., n,$$

and

$$\eta_i(\lambda) = \xi_i - b_i(\lambda), \quad i = 1, ..., n,$$

we obtain the following decomposition:

$$S_k = T_k(\lambda) + Y_k(\lambda), \quad k = 1, ..., n, \tag{39}$$

where

$$T_k(\lambda) = \sum_{i=1}^{k} b_i(\lambda) \qquad \text{and} \qquad Y_k(\lambda) = \sum_{i=1}^{k}\eta_i(\lambda).$$

In the following, we give some lower and upper bounds of $T_n(\lambda)$, which will be used in the proofs of theorems.

Lemma 4.1. *For all* $0 \leq \lambda < \varepsilon^{-1}$,

$$(1-2.4\lambda\varepsilon)\lambda\sigma^2 \leq \frac{(1-1.5\lambda\varepsilon)(1-\lambda\varepsilon)}{1-\lambda\varepsilon+6\lambda^2\varepsilon^2}\lambda\sigma^2 \leq T_n(\lambda) \leq \frac{1-0.5\lambda\varepsilon}{(1-\lambda\varepsilon)^2}\lambda\sigma^2.$$

Proof. Since $\mathbf{E}[\xi_i]=0$, by Jensen's inequality, we have $\mathbf{E}[e^{\lambda\xi_i}] \geq 1$. Noting that

$$\mathbf{E}[\xi_i e^{\lambda\xi_i}] = \mathbf{E}[\xi_i(e^{\lambda\xi_i}-1)] \geq 0, \qquad \lambda \geq 0,$$

by Taylor's expansion of e^x, we get

$$\begin{aligned} T_n(\lambda) &\leq \sum_{i=1}^n \mathbf{E}[\xi_i e^{\lambda\xi_i}] \\ &= \lambda\sigma^2 + \sum_{i=1}^n \sum_{k=2}^{+\infty} \frac{\lambda^k}{k!}\mathbf{E}[\xi_i^{k+1}]. \end{aligned} \tag{40}$$

Using Bernstein's condition (2), we obtain, for all $0 \leq \lambda < \varepsilon^{-1}$,

$$\begin{aligned} \sum_{i=1}^n \sum_{k=2}^{+\infty} \frac{\lambda^k}{k!}|\mathbf{E}[\xi_i^{k+1}]| &\leq \frac{1}{2}\lambda^2\sigma^2\varepsilon \sum_{k=2}^{+\infty}(k+1)(\lambda\varepsilon)^{k-2} \\ &= \frac{3-2\lambda\varepsilon}{2(1-\lambda\varepsilon)^2}\lambda^2\sigma^2\varepsilon. \end{aligned} \tag{41}$$

Combining (40) and (41), we get the desired upper bound of $T_n(\lambda)$. By Jensen's inequality and Bernstein's condition (2),

$$(\mathbf{E}[\xi_i^2])^2 \leq \mathbf{E}[\xi_i^4] \leq 12\varepsilon^2\mathbf{E}[\xi_i^2],$$

from which we get

$$\mathbf{E}[\xi_i^2] \leq 12\varepsilon^2.$$

Using again Bernstein's condition (2), we have, for all $0 \leq \lambda < \varepsilon^{-1}$,

$$\begin{aligned} \mathbf{E}[e^{\lambda\xi_i}] &\leq 1 + \sum_{k=2}^{+\infty}\frac{\lambda^k}{k!}|\mathbf{E}[\xi_i^k]| \\ &\leq 1 + \frac{\lambda^2\mathbf{E}[\xi_i^2]}{2(1-\lambda\varepsilon)} \\ &\leq 1 + \frac{6\lambda^2\varepsilon^2}{1-\lambda\varepsilon} \\ &= \frac{1-\lambda\varepsilon+6\lambda^2\varepsilon^2}{1-\lambda\varepsilon}. \end{aligned} \tag{42}$$

Notice that $g(t) = e^t - (1+t+\frac{1}{2}t^2)$ satisfies that $g(t) > 0$ if $t > 0$ and $g(t) < 0$ if $t < 0$, which leads to $tg(t) \geq 0$ for all $t \in \mathbf{R}$. That is, $te^t \geq t(1+t+\frac{1}{2}t^2)$ for all $t \in \mathbf{R}$. Therefore, for all $0 \leq \lambda < \varepsilon^{-1}$,

$$\xi_i e^{\lambda\xi_i} \geq \xi_i\left(1+\lambda\xi_i+\frac{\lambda^2\xi_i^2}{2}\right).$$

Taking expectation, we get

$$\mathbf{E}[\xi_i e^{\lambda\xi_i}] \geq \lambda\mathbf{E}[\xi_i^2] + \frac{\lambda^2}{2}\mathbf{E}[\xi_i^3] \geq \lambda\mathbf{E}[\xi_i^2] - \frac{\lambda^2}{2}\frac{1}{2}3!\varepsilon\mathbf{E}[\xi_i^2] = (1-1.5\lambda\varepsilon)\lambda\mathbf{E}[\xi_i^2],$$

from which, it follows that

$$\sum_{i=1}^{n} \mathbf{E}[\xi_i e^{\lambda \xi_i}] \quad \geq \quad (1 - 1.5\lambda\varepsilon)\lambda\sigma^2. \tag{43}$$

Combining (42) and (43), we obtain the following lower bound of $T_n(\lambda)$: for all $0 \leq \lambda < \varepsilon^{-1}$,

$$\begin{aligned} T_n(\lambda) \quad &\geq \quad \sum_{i=1}^{n} \frac{\mathbf{E}[\xi_i e^{\lambda \xi_i}]}{\mathbf{E}[e^{\lambda \xi_i}]} \\ &\geq \quad \frac{(1 - 1.5\lambda\varepsilon)(1 - \lambda\varepsilon)}{1 - \lambda\varepsilon + 6\lambda^2\varepsilon^2}\lambda\sigma^2 \\ &\geq \quad (1 - 2.4\lambda\varepsilon)\lambda\sigma^2. \end{aligned} \tag{44}$$

This completes the proof of Lemma 4.1. □

We now consider the following *cumulant* function

$$\Psi_n(\lambda) = \sum_{i=1}^{n} \log \mathbf{E}[e^{\lambda \xi_i}], \qquad 0 \leq \lambda < \varepsilon^{-1}. \tag{45}$$

We have the following elementary bound for $\Psi_n(\lambda)$.

Lemma 4.2. *For all* $0 \leq \lambda < \varepsilon^{-1}$,

$$\Psi_n(\lambda) \leq n \log\left(1 + \frac{\lambda^2\sigma^2}{2n(1 - \lambda\varepsilon)}\right) \leq \frac{\lambda^2\sigma^2}{2(1 - \lambda\varepsilon)}$$

and

$$-\lambda T_n(\lambda) + \Psi_n(\lambda) \quad \geq \quad -\frac{\lambda^2\sigma^2}{2(1 - \lambda\varepsilon)^6}.$$

Proof. By Bernstein's condition (2), it is easy to see that, for all $0 \leq \lambda < \varepsilon^{-1}$,

$$\mathbf{E}[e^{\lambda \xi_i}] \quad = \quad 1 + \sum_{k=2}^{+\infty} \frac{\lambda^k}{k!} \mathbf{E}[\xi_i^k] \leq 1 + \frac{\lambda^2}{2} \mathbf{E}[\xi_i^2] \sum_{k=2}^{\infty} (\lambda\varepsilon)^{k-2} = 1 + \frac{\lambda^2 \mathbf{E}[\xi_i^2]}{2(1 - \lambda\varepsilon)}.$$

Then, we have

$$\Psi_n(\lambda) \quad \leq \quad \sum_{i=1}^{n} \log\left(1 + \frac{\lambda^2 \mathbf{E}[\xi_i^2]}{2(1 - \lambda\varepsilon)}\right). \tag{46}$$

Using the fact $\log(1 + t)$ is concave in $t \geq 0$ and $\log(1 + t) \leq t$, we get the first assertion of the lemma. Since $\Psi_n(0) = 0$ and $\Psi_n'(\lambda) = T_n(\lambda)$, by Lemma 4.1, for all $0 \leq \lambda < \varepsilon^{-1}$,

$$\Psi_n(\lambda) = \int_0^{\lambda} T_n(t)dt \geq \int_0^{\lambda} t(1 - 2.4t\varepsilon)\sigma^2 dt = \frac{\lambda^2\sigma^2}{2}(1 - 1.6\lambda\varepsilon).$$

Therefore, using again Lemma 4.1, we see that

$$\begin{aligned} -\lambda T_n(\lambda) + \Psi_n(\lambda) \quad &\geq \quad -\frac{1 - 0.5\lambda\varepsilon}{(1 - \lambda\varepsilon)^2}\lambda^2\sigma^2 + \frac{\lambda^2\sigma^2}{2}(1 - 1.6\lambda\varepsilon) \\ &\geq \quad -\frac{\lambda^2\sigma^2}{2(1 - \lambda\varepsilon)^6}, \end{aligned}$$

which completes the proof of the second assertion of the lemma. $\square$

Denote $\overline{\sigma}^2(\lambda) = \mathbf{E}_\lambda[Y_n^2(\lambda)]$. By the relation between $\mathbf{E}$ and $\mathbf{E}_\lambda$, we have

$$\overline{\sigma}^2(\lambda) = \sum_{i=1}^{n} \left(\frac{\mathbf{E}[\xi_i^2 e^{\lambda\xi_i}]}{\mathbf{E}[e^{\lambda\xi_i}]} - \frac{(\mathbf{E}[\xi_i e^{\lambda\xi_i}])^2}{(\mathbf{E}[e^{\lambda\xi_i}])^2} \right), \qquad 0 \leq \lambda < \varepsilon^{-1}.$$

Lemma 4.3. *For all* $0 \leq \lambda < \varepsilon^{-1}$,

$$\frac{(1-\lambda\varepsilon)^2(1-3\lambda\varepsilon)}{(1-\lambda\varepsilon+6\lambda^2\varepsilon^2)^2}\sigma^2 \leq \overline{\sigma}^2(\lambda) \leq \frac{\sigma^2}{(1-\lambda\varepsilon)^3}. \tag{47}$$

Proof. Denote $f(\lambda) = \mathbf{E}[\xi_i^2 e^{\lambda\xi_i}]\mathbf{E}[e^{\lambda\xi_i}] - (\mathbf{E}[\xi_i e^{\lambda\xi_i}])^2$. Then,

$$f'(0) = \mathbf{E}[\xi_i^3] \quad \text{and} \quad f''(\lambda) \;=\; \mathbf{E}[\xi_i^4 e^{\lambda\xi_i}]\mathbf{E}[e^{\lambda\xi_i}] - (\mathbf{E}[\xi_i^2 e^{\lambda\xi_i}])^2 \geq 0.$$

Thus,

$$f(\lambda) \geq f(0) + f'(0)\lambda = \mathbf{E}[\xi_i^2] + \lambda\mathbf{E}[\xi_i^3]. \tag{48}$$

Using (48), (42) and Bernstein's condition (2), we have, for all $0 \leq \lambda < \varepsilon^{-1}$,

$$\begin{aligned}
\mathbf{E}_\lambda[\eta_i^2] &= \frac{\mathbf{E}[\xi_i^2 e^{\lambda\xi_i}]\mathbf{E}[e^{\lambda\xi_i}] - (\mathbf{E}[\xi_i e^{\lambda\xi_i}])^2}{(\mathbf{E}[e^{\lambda\xi_i}])^2} \\
&\geq \frac{\mathbf{E}[\xi_i^2] + \lambda\mathbf{E}[\xi_i^3]}{(\mathbf{E}[e^{\lambda\xi_i}])^2} \\
&\geq \left(\frac{1-\lambda\varepsilon}{1-\lambda\varepsilon+6\lambda^2\varepsilon^2}\right)^2 (\mathbf{E}[\xi_i^2] + \lambda\mathbf{E}[\xi_i^3]) \\
&\geq \frac{(1-\lambda\varepsilon)^2(1-3\lambda\varepsilon)}{(1-\lambda\varepsilon+6\lambda^2\varepsilon^2)^2}\mathbf{E}[\xi_i^2].
\end{aligned}$$

Therefore

$$\overline{\sigma}^2(\lambda) \geq \frac{(1-\lambda\varepsilon)^2(1-3\lambda\varepsilon)}{(1-\lambda\varepsilon+6\lambda^2\varepsilon^2)^2}\sigma^2.$$

Using Taylor's expansion of e^x and Bernstein's condition (2) again, we obtain

$$\overline{\sigma}^2(\lambda) \leq \sum_{i=1}^{n} \mathbf{E}[\xi_i^2 e^{\lambda\xi_i}] \leq \frac{\sigma^2}{(1-\lambda\varepsilon)^3}.$$

This completes the proof of Lemma 4.3. $\square$

For the r.v. $Y_n(\lambda)$ with $0 \leq \lambda < \varepsilon^{-1}$, we have the following result on the rate of convergence to the standard normal law.

Lemma 4.4. *For all* $0 \leq \lambda < \varepsilon^{-1}$,

$$\sup_{y\in\mathbf{R}} \left| \mathbf{P}_\lambda\left(\frac{Y_n(\lambda)}{\overline{\sigma}(\lambda)} \leq y\right) - \Phi(y) \right| \leq 13.44 \frac{\sigma^2\varepsilon}{\overline{\sigma}^3(\lambda)(1-\lambda\varepsilon)^4}.$$

Proof. Since $Y_n(\lambda) = \sum_{i=1}^n \eta_i(\lambda)$ is the sum of independent and centered (respect to $\mathbf{P}_\lambda$) r.v.'s $\eta_i(\lambda)$, using standard results on the rate of convergence in the central limit theorem (cf. e.g. Petrov [34], p. 115) we get, for $0 \leq \lambda < \varepsilon^{-1}$,

$$\sup_{y\in\mathbf{R}} \left| \mathbf{P}_\lambda\left(\frac{Y_n(\lambda)}{\overline{\sigma}(\lambda)} \leq y\right) - \Phi(y) \right| \leq C_1 \frac{1}{\overline{\sigma}^3(\lambda)} \sum_{i=1}^{n} \mathbf{E}_\lambda[|\eta_i|^3],$$

where $C_1 > 0$ is an absolute constant. For $0 \leq \lambda < \varepsilon^{-1}$, using Bernstein's condition, we have

$$\begin{aligned}\sum_{i=1}^n \mathbf{E}_\lambda[|\eta_i|^3] &\leq 4\sum_{i=1}^n \mathbf{E}_\lambda[|\xi_i|^3 + (\mathbf{E}_\lambda[|\xi_i|])^3] \\ &\leq 8\sum_{i=1}^n \mathbf{E}_\lambda[|\xi_i|^3] \\ &\leq 8\sum_{i=1}^n \mathbf{E}[|\xi_i|^3 \exp\{|\lambda\xi_i|\}] \\ &\leq 8\sum_{i=1}^n \mathbf{E}\Big[\sum_{j=0}^\infty \frac{\lambda^j}{j!}|\xi_i|^{3+j}\Big] \\ &\leq 4\sigma^2\varepsilon\sum_{j=0}^\infty (j+3)(j+2)(j+1)(\lambda\varepsilon)^j.\end{aligned}$$

As

$$\sum_{j=0}^\infty (j+3)(j+2)(j+1)x^j = \frac{d^3}{dx^3}\sum_{j=0}^\infty x^j = \frac{6}{(1-x)^4}, \quad |x| < 1,$$

we obtain, for $0 \leq \lambda < \varepsilon^{-1}$,

$$\sum_{i=1}^n \mathbf{E}_\lambda[|\eta_i|^3] \leq 24\frac{\sigma^2\varepsilon}{(1-\lambda\varepsilon)^4}.$$

Therefore, we have, for $0 \leq \lambda < \varepsilon^{-1}$,

$$\begin{aligned}\sup_{y\in\mathbf{R}}\left|\mathbf{P}_\lambda\left(\frac{Y_n(\lambda)}{\overline{\sigma}(\lambda)} \leq y\right) - \Phi(y)\right| &\leq 24C_1\,\frac{\sigma^2\varepsilon}{\overline{\sigma}^3(\lambda)(1-\lambda\varepsilon)^4} \\ &\leq 13.44\frac{\sigma^2\varepsilon}{\overline{\sigma}^3(\lambda)(1-\lambda\varepsilon)^4},\end{aligned}$$

where the last step holds as $C_1 \leq 0.56$ (cf. Shevtsova [42]). $\square$

Using Lemma 4.4, we easily obtain the following lemma.

Lemma 4.5. *For all* $0 \leq \lambda \leq 0.1\varepsilon^{-1}$,

$$\sup_{y\in\mathbf{R}}\left|\mathbf{P}_\lambda\left(Y_n(\lambda) \leq \frac{y\sigma}{1-\lambda\varepsilon}\right) - \Phi(y)\right| \leq 1.07\lambda\varepsilon + 42.45\frac{\varepsilon}{\sigma}.$$

Proof. Using Lemma 4.3, we have, for all $0 \leq \lambda < \frac{1}{3}\varepsilon^{-1}$,

$$\sqrt{1-\lambda\varepsilon} \leq \frac{\sigma}{\overline{\sigma}(\lambda)(1-\lambda\varepsilon)} \leq \frac{1-\lambda\varepsilon+6\lambda^2\varepsilon^2}{(1-\lambda\varepsilon)^2\sqrt{1-3\lambda\varepsilon}}. \tag{49}$$

It is easy to see that

$$\begin{aligned}&\left|\mathbf{P}_\lambda\left(Y_n(\lambda) \leq \frac{y\sigma}{1-\lambda\varepsilon}\right) - \Phi(y)\right| \\ \leq\ &\left|\mathbf{P}_\lambda\left(\frac{Y_n(\lambda)}{\overline{\sigma}(\lambda)} \leq \frac{y\sigma}{\overline{\sigma}(\lambda)(1-\lambda\varepsilon)}\right) - \Phi\left(\frac{y\sigma}{\overline{\sigma}(\lambda)(1-\lambda\varepsilon)}\right)\right| \\ &+\left|\Phi\left(\frac{y\sigma}{\overline{\sigma}(\lambda)(1-\lambda\varepsilon)}\right) - \Phi(y)\right| \\ =:\ &I_1 + I_2.\end{aligned}$$

By Lemma 4.4 and (49), we get, for all $0 \leq \lambda < \frac{1}{3}\varepsilon^{-1}$,

$$I_1 \ \leq\ 13.44\ \frac{\sigma^2 \varepsilon}{\overline{\sigma}^3(\lambda)(1-\lambda\varepsilon)^4}\ \leq\ 13.44 R(\lambda\varepsilon)\frac{\varepsilon}{\sigma}.$$

Using Taylor's expansion and (49), we obtain, for all $0 \leq \lambda < \frac{1}{3}\varepsilon^{-1}$,

$$\begin{aligned} I_2 &\leq \frac{1}{\sqrt{2\pi}} y e^{-\frac{y^2(1-\lambda\varepsilon)}{2}} \left| \frac{\sigma}{\overline{\sigma}(\lambda)(1-\lambda\varepsilon)} - 1 \right| \\ &\leq \frac{1}{\sqrt{2\pi}} y e^{-\frac{y^2(1-\lambda\varepsilon)}{2}} \left(\left| \frac{1-\lambda\varepsilon+6\lambda^2\varepsilon^2}{(1-\lambda\varepsilon)^2\sqrt{1-3\lambda\varepsilon}} - 1 \right| \vee \left| 1-\sqrt{1-\lambda\varepsilon} \right| \right) \\ &\leq \frac{1}{\sqrt{2e\pi(1-\lambda\varepsilon)}} \left| \frac{1-\lambda\varepsilon+6\lambda^2\varepsilon^2}{(1-\lambda\varepsilon)^2\sqrt{1-3\lambda\varepsilon}} - 1 \right|. \end{aligned}$$

By simple calculations, we obtain, for all $0 \leq \lambda \leq 0.1\varepsilon^{-1}$,

$$\left| \mathbf{P}_\lambda \left(Y_n(\lambda) \leq \frac{y\sigma}{1-\lambda\varepsilon} \right) - \Phi(y) \right| \leq 1.07\lambda\varepsilon + 42.45\frac{\varepsilon}{\sigma}.$$

This completes the proof of Lemma 4.5. □

5. Proofs of Theorems 2.1-2.2

In this section, we give upper bounds for $\mathbf{P}(S_n > x\sigma)$. For all $x \geq 0$ and $0 \leq \lambda < \varepsilon^{-1}$, by (38) and (39), we have:

$$\begin{aligned} \mathbf{P}(S_n > x\sigma) &= \mathbf{E}_\lambda[Z_n(\lambda)^{-1}\mathbf{1}_{\{S_n > x\sigma\}}] \\ &= \mathbf{E}_\lambda[e^{-\lambda S_n + \Psi_n(\lambda)}\mathbf{1}_{\{S_n > x\sigma\}}] \\ &= \mathbf{E}_\lambda[e^{-\lambda T_n(\lambda) + \Psi_n(\lambda) - \lambda Y_n(\lambda)}\mathbf{1}_{\{Y_n(\lambda)+T_n(\lambda)-x\sigma>0\}}]. \end{aligned} \tag{50}$$

Setting $U_n(\lambda) = \lambda(Y_n(\lambda) + T_n(\lambda) - x\sigma)$, we get

$$\mathbf{P}(S_n > x\sigma) \ = \ e^{-\lambda x\sigma + \Psi_n(\lambda)}\mathbf{E}_\lambda[e^{-U_n(\lambda)}\mathbf{1}_{\{U_n(\lambda)>0\}}].$$

Then, we deduce, for all $x \geq 0$ and $0 \leq \lambda < \varepsilon^{-1}$,

$$\mathbf{P}(S_n > x\sigma) \ = \ e^{-\lambda x\sigma + \Psi_n(\lambda)} \int_0^\infty e^{-t}\mathbf{P}_\lambda(0 < U_n(\lambda) \leq t)dt. \tag{51}$$

In the sequel, denote by $N(0,1)$ a standard normal r.v.

5.1. Proof of Theorem 2.1

From (51), using Lemma 4.2, we obtain, for all $x \geq 0$ and $0 \leq \lambda < \varepsilon^{-1}$,

$$\mathbf{P}(S_n > x\sigma) \ \leq \ e^{-\lambda x\sigma + \frac{\lambda^2\sigma^2}{2(1-\lambda\varepsilon)}} \int_0^\infty e^{-t}\mathbf{P}_\lambda(0 < U_n(\lambda) \leq t)dt. \tag{52}$$

For any $x \geq 0$ and $\beta \in [0, 0.5)$, let $\overline{\lambda} = \overline{\lambda}(x) \in [0, \varepsilon^{-1})$ be the unique solution of the equation

$$\frac{\lambda - \beta\lambda^2\varepsilon}{(1-\lambda\varepsilon)^2} = \frac{x}{\sigma}.$$

This definition and Lemma 4.1 implies that

$$\overline{\lambda} = \frac{2x/\sigma}{1 + 2x\varepsilon/\sigma + \sqrt{1 + 4(1-\beta)x\varepsilon/\sigma}} \qquad \text{and} \qquad T_n(\overline{\lambda}) \leq x\sigma. \tag{53}$$

Using (52) with $\lambda = \overline{\lambda}$, we get

$$\mathbf{P}(S_n > x\sigma) \leq e^{-\frac{1}{2}\left(1+(1-2\beta)\overline{\lambda}\varepsilon\right)\widetilde{x}^{\,2}} \int_0^\infty e^{-t}\mathbf{P}_{\overline{\lambda}}(0 < U_n(\overline{\lambda}) \leq t)dt, \tag{54}$$

where

$$\widetilde{x} = \frac{\overline{\lambda}\sigma}{1 - \overline{\lambda}\varepsilon}.$$

By (53) and Lemma 4.5, we have, for $0 \leq \overline{\lambda} \leq 0.1\varepsilon^{-1}$,

$$\begin{aligned}
& \int_0^\infty e^{-t}\mathbf{P}_{\overline{\lambda}}(0 < U_n(\overline{\lambda}) \leq t)dt \\
= & \int_0^\infty e^{-y\widetilde{x}}\mathbf{P}_{\overline{\lambda}}\left(0 < U_n(\overline{\lambda}) \leq y\widetilde{x}\right)\widetilde{x}dy \\
\leq & \int_0^\infty e^{-y\widetilde{x}}\mathbf{P}\left(0 < N(0,1) \leq y\right)\widetilde{x}dy + 2\left(1.07\overline{\lambda}\varepsilon + 42.45\frac{\varepsilon}{\sigma}\right) \\
= & \ M\left(\widetilde{x}\right) + 2.14\overline{\lambda}\varepsilon + 84.9\frac{\varepsilon}{\sigma}.
\end{aligned} \tag{55}$$

Since $\int_0^\infty e^{-t}\mathbf{P}_{\overline{\lambda}}(0 < U_n(\overline{\lambda}) \leq t)dt \leq 1$ and $M^{-1}(t) \leq \sqrt{2\pi}\,(1+t)$ for $t \geq 0$ (cf. (26)), combining (54) and (55), we deduce, for all $x \geq 0$,

$$\begin{aligned}
& \mathbf{P}(S_n > x\sigma) \\
\leq & \ e^{-\frac{1}{2}(1-2\beta)\overline{\lambda}\varepsilon\widetilde{x}^{\,2} - \frac{1}{2}\widetilde{x}^{\,2}}\mathbf{1}_{\{\overline{\lambda}\varepsilon > 0.1\}} \\
& + e^{-\frac{1}{2}(1-2\beta)\overline{\lambda}\varepsilon\widetilde{x}^{\,2}}\left[1 - \Phi\left(\widetilde{x}\right) + e^{-\frac{1}{2}\widetilde{x}^{\,2}}\left(2.14\overline{\lambda}\varepsilon + 84.9\frac{\varepsilon}{\sigma}\right)\right]\mathbf{1}_{\{\overline{\lambda}\varepsilon \leq 0.1\}} \\
\leq & \ \left(1 - \Phi\left(\widetilde{x}\right)\right)\left(I_{11} + I_{12}\right),
\end{aligned} \tag{56}$$

with

$$I_{11} = \exp\left\{-\frac{1}{2}(1-2\beta)\overline{\lambda}\varepsilon\widetilde{x}^2\right\}\left[\sqrt{2\pi}\,(1+\widetilde{x})\right]\mathbf{1}_{\{\overline{\lambda}\varepsilon > 0.1\}} \tag{57}$$

and

$$I_{12} \ = \ e^{-\frac{1}{2}(1-2\beta)\overline{\lambda}\varepsilon\widetilde{x}^{\,2}}\left[1 + \sqrt{2\pi}\,(1+\widetilde{x})\left(2.14\overline{\lambda}\varepsilon + 84.9\frac{\varepsilon}{\sigma}\right)\right]\mathbf{1}_{\{\overline{\lambda}\varepsilon \leq 0.1\}}.$$

Now we shall give estimates for I_{11} and I_{12}. If $\overline{\lambda}\varepsilon > 0.1$, then $I_{12} = 0$ and

$$I_{11} \ \leq \ \exp\left\{-0.1(1-2\beta)\frac{\widetilde{x}^2}{2}\right\}\left[\sqrt{2\pi}\,(1+\widetilde{x})\right]. \tag{58}$$

By a simple calculation, $I_{11} \leq 1$ provided that $\widetilde{x} \geq \frac{8}{1-2\beta}$ (note that $\beta \in [0, 0.5)$). For $0 \leq \widetilde{x} < \frac{8}{1-2\beta}$, we get $\overline{\lambda}\sigma = \widetilde{x}(1 - \overline{\lambda}\varepsilon) < \frac{8}{1-2\beta}(1 - 0.1) = \frac{7.2}{1-2\beta}$. Then, using $10\overline{\lambda}\varepsilon > 1$, we obtain

$$\begin{aligned}
I_{11} \ \leq & \ 1 + \sqrt{2\pi}\,(1+\widetilde{x}) \\
\leq & \ 1 + 10\sqrt{2\pi}\,(1+\widetilde{x})\,\overline{\lambda}\sigma\frac{\varepsilon}{\sigma} \\
\leq & \ 1 + \frac{180.48}{1-2\beta}\,(1+\widetilde{x})\,\frac{\varepsilon}{\sigma}.
\end{aligned}$$

If $0 \le \overline{\lambda}\varepsilon \le 0.1$, we have $I_{11} = 0$. Since

$$\begin{aligned} & 1+\sqrt{2\pi}\,(1+\widetilde{x})\left(2.14\overline{\lambda}\varepsilon+84.9\frac{\varepsilon}{\sigma}\right) \\ \le\ & \left(1+2.14\sqrt{2\pi}\,(1+\widetilde{x})\,\overline{\lambda}\varepsilon\right)\left(1+84.9\sqrt{2\pi}\,(1+\widetilde{x})\,\frac{\varepsilon}{\sigma}\right) \\ =\ & J_1 J_2, \end{aligned}$$

it follows that $I_{12} \le \exp\left\{-\frac{1}{2}(1-2\beta)\overline{\lambda}\varepsilon\widetilde{x}^2\right\} J_1 J_2$. Using the inequality $1+x \le e^x$, we deduce

$$I_{12} \le \exp\left\{-\overline{\lambda}\varepsilon\left((1-2\beta)\frac{\widetilde{x}^2}{2} - 2.14\sqrt{2\pi}\,(1+\widetilde{x})\right)\right\} J_2.$$

If $\widetilde{x} \ge \frac{11.65}{1-2\beta}$, we see that $\frac{1}{2}(1-2\beta)\widetilde{x}^2 - 2.14\sqrt{2\pi}\,(1+\widetilde{x}) \ge 0$, so $I_{12} \le J_2$. For $0 \le \widetilde{x} < \frac{11.65}{1-2\beta}$, we get $\overline{\lambda}\sigma = \widetilde{x}(1-\overline{\lambda}\varepsilon) < \frac{11.65}{1-2\beta}$. Then

$$\begin{aligned} I_{12} &\le 1+\sqrt{2\pi}\,(1+\widetilde{x})\left(2.14\overline{\lambda}\varepsilon+84.9\frac{\varepsilon}{\sigma}\right) \\ &< 1+\sqrt{2\pi}\,(1+\widetilde{x})\left(2.14\frac{11.65}{1-2\beta}+84.9\right)\frac{\varepsilon}{\sigma} \\ &\le 1+\left(\frac{62.493}{1-2\beta}+212.813\right)(1+\widetilde{x})\,\frac{\varepsilon}{\sigma}. \end{aligned}$$

Hence, whenever $0 \le \overline{\lambda}\varepsilon < 1$, we have

$$I_{11}+I_{12} \le 1+\left(\left(\frac{62.493}{1-2\beta}+212.813\right)\vee\frac{180.48}{1-2\beta}\right)(1+\widetilde{x})\,\frac{\varepsilon}{\sigma}. \tag{59}$$

Therefore, substituting $\overline{\lambda}$ from (53) in the expression of $\widetilde{x} = \frac{\overline{\lambda}\sigma}{1-\overline{\lambda}\varepsilon}$ and replacing $1-2\beta$ by δ, we obtain inequality (10) in Theorem 2.1 from (56) and (59).

5.2. Proof of Theorem 2.2

For any $x \ge 0$, let $\overline{\lambda} = \overline{\lambda}(x) \in [0, \varepsilon^{-1})$ be the unique solution of the equation

$$\frac{\lambda - 0.5\lambda^2\varepsilon}{(1-\lambda\varepsilon)^2} = \frac{x}{\sigma}. \tag{60}$$

By Lemma 4.1, it follows that

$$\overline{\lambda} = \frac{2x/\sigma}{1+2x\varepsilon/\sigma+\sqrt{1+2x\varepsilon/\sigma}} \quad \text{and} \quad T_n(\overline{\lambda}) \le x\sigma. \tag{61}$$

Using Lemma 4.4 and $T_n(\overline{\lambda}) \le x\sigma$, we have, for all $0 \le \overline{\lambda} < \varepsilon^{-1}$,

$$\begin{aligned} & \int_0^\infty e^{-t}\mathbf{P}_{\overline{\lambda}}(0 < U_n(\overline{\lambda}) \le t)dt \\ =\ & \int_0^\infty e^{-y\overline{\lambda}\overline{\sigma}(\overline{\lambda})}\mathbf{P}_{\overline{\lambda}}\left(0 < U_n(\overline{\lambda}) \le y\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right)\overline{\lambda}\overline{\sigma}(\overline{\lambda})dy \\ \le\ & \int_0^\infty e^{-y\overline{\lambda}\overline{\sigma}(\overline{\lambda})}\mathbf{P}\left(0 < N(0,1) \le y\right)\overline{\lambda}\overline{\sigma}(\overline{\lambda})dy + 26.88\,\frac{\sigma^2\varepsilon}{\overline{\sigma}^3(\overline{\lambda})(1-\overline{\lambda}\varepsilon)^4} \\ \le\ & \int_0^\infty e^{-y\overline{\lambda}\overline{\sigma}(\overline{\lambda})}d\Phi(y) + 26.88\,\frac{\sigma^2\varepsilon}{\overline{\sigma}^3(\overline{\lambda})(1-\overline{\lambda}\varepsilon)^4} \\ =\ & F := M\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) + 26.88\,\frac{\sigma^2\varepsilon}{\overline{\sigma}^3(\overline{\lambda})(1-\overline{\lambda}\varepsilon)^4}. \end{aligned} \tag{62}$$

Using $\lambda = \overline{\lambda}$ and $\int_0^\infty e^{-t}\mathbf{P}_{\overline{\lambda}}(0 < U_n(\overline{\lambda}) \leq t)dt \leq 1$, from (51) and (62), we obtain

$$\mathbf{P}(S_n > x\sigma) \quad \leq \quad [F \wedge 1] \times \exp\left\{-\overline{\lambda}x\sigma + \Psi_n(\overline{\lambda})\right\}.$$

By Lemma 4.2, inequality (63) implies that

$$\mathbf{P}(S_n > x\sigma) \quad \leq \quad [F \wedge 1] \times \exp\left\{-\overline{\lambda}x\sigma + n\log\left(1 + \frac{\overline{\lambda}^2\sigma^2}{2n(1-\overline{\lambda}\varepsilon)}\right)\right\}.$$

Substituting $\overline{\lambda}$ from (61) in the previous exponential function, we get

$$\mathbf{P}(S_n > x\sigma) \quad \leq \quad [F \wedge 1] \times B_n\left(x, \frac{\varepsilon}{\sigma}\right). \tag{63}$$

Next, we give an estimation of F. Since $M(t)$ is decreasing in $t \geq 0$ and $|M'(t)| \leq \frac{1}{\sqrt{\pi}\,t^2}, t > 0$, it follows that

$$M\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) - M(x) \quad \leq \quad \frac{1}{\sqrt{\pi}}\frac{1}{\overline{\lambda}^2\overline{\sigma}^2(\overline{\lambda})}\left(x - \overline{\lambda}\overline{\sigma}(\overline{\lambda})\right)^+.$$

Using Lemma 4.3, by a simple calculation, we deduce

$$\begin{aligned}
& M\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) - M(x) \\
\leq \quad & \frac{1}{\sqrt{\pi}}\frac{\overline{\lambda}\sigma}{\overline{\lambda}^2\overline{\sigma}^2(\overline{\lambda})}\left(\frac{1-0.5\overline{\lambda}\varepsilon}{(1-\overline{\lambda}\varepsilon)^2} - \frac{(1-\overline{\lambda}\varepsilon)\sqrt{(1-3\overline{\lambda}\varepsilon)^+}}{1-\overline{\lambda}\varepsilon+6\overline{\lambda}^2\varepsilon^2}\right) \\
\leq \quad & \frac{(1-0.5\overline{\lambda}\varepsilon)(1-\overline{\lambda}\varepsilon+6\overline{\lambda}^2\varepsilon^2) - (1-\overline{\lambda}\varepsilon)\sqrt{(1-3\overline{\lambda}\varepsilon)^+}}{\sqrt{\pi}\;\overline{\lambda}\varepsilon(1-\overline{\lambda}\varepsilon)^4(1-3\overline{\lambda}\varepsilon)^+/(1-\overline{\lambda}\varepsilon+6\overline{\lambda}^2\varepsilon^2)}\,\frac{\varepsilon}{\sigma} \\
\leq \quad & 1.11R\left(\overline{\lambda}\varepsilon\right)\frac{\varepsilon}{\sigma}.
\end{aligned} \tag{64}$$

By Lemma 4.3, it is easy to see that

$$26.88\,\frac{\sigma^2\varepsilon}{\overline{\sigma}^3(\overline{\lambda})(1-\overline{\lambda}\varepsilon)^4} \leq 26.88R\left(\overline{\lambda}\varepsilon\right)\frac{\varepsilon}{\sigma}. \tag{65}$$

Hence, it follows from (62), (64) and (65) that

$$F \;\leq\; M(x) + 27.99R\left(\overline{\lambda}\varepsilon\right)\frac{\varepsilon}{\sigma}. \tag{66}$$

Implementing (66) into (63) and using $\overline{\lambda}\varepsilon \leq x\frac{\varepsilon}{\sigma}$, we obtain inequality (13).

6. Proof of Theorem 2.3

In this section, we give a lower bound for $\mathbf{P}(S_n > x\sigma)$. From Lemma 4.2 and (50), it follows that, for all $0 \leq \lambda < \varepsilon^{-1}$,

$$\mathbf{P}(S_n > x\sigma) \quad \geq \quad \exp\left\{-\frac{\lambda^2\sigma^2}{2(1-\lambda\varepsilon)^6}\right\}\mathbf{E}_\lambda[e^{-\lambda Y_n(\lambda)}\mathbf{1}_{\{Y_n(\lambda)+T_n(\lambda)-x\sigma>0\}}].$$

Let $\overline{\lambda} = \overline{\lambda}(x) \in [0, \varepsilon^{-1}/4.8]$ be the unique solution of the equation

$$\lambda(1-2.4\lambda\varepsilon)\sigma^2 = x\sigma. \tag{67}$$

This definition and Lemma 4.1 implies that, for all $0 \leq x \leq \sigma/(9.6\varepsilon)$,

$$\overline{\lambda} = \frac{2x/\sigma}{1+\sqrt{1-9.6x\varepsilon/\sigma}} \qquad \text{and} \qquad x\sigma \leq T_n(\overline{\lambda}). \tag{68}$$

Therefore,

$$\mathbf{P}(S_n > x\sigma) \quad \geq \quad \exp\left\{-\frac{\lambda^2\sigma^2}{2(1-\lambda\varepsilon)^6}\right\}\mathbf{E}_\lambda[e^{-\lambda Y_n(\lambda)}\mathbf{1}_{\{Y_n(\lambda)>0\}}].$$

Setting $V_n(\overline{\lambda}) = \overline{\lambda}Y_n(\overline{\lambda})$, we get

$$\mathbf{P}(S_n > x\sigma) \quad \geq \quad \exp\left\{-\frac{\check{x}^2}{2}\right\}\int_0^\infty e^{-t}\mathbf{P}_{\overline{\lambda}}(0 < V_n(\overline{\lambda}) \leq t)dt, \tag{69}$$

where $\check{x} = \frac{\overline{\lambda}\sigma}{(1-\overline{\lambda}\varepsilon)^3}$. By Lemma 4.4 and an argument similar to that used to prove (62), it is easy to see that

$$\int_0^\infty e^{-t}\mathbf{P}_{\overline{\lambda}}(0 < V_n(\overline{\lambda}) \leq t)dt \quad \geq \quad M\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) - G,$$

where $G = 26.88\, \frac{\sigma^2\varepsilon}{\overline{\sigma}^3(\overline{\lambda})(1-\overline{\lambda}\varepsilon)^4}$. Since $M(t)$ is decreasing in $t \geq 0$ and $\overline{\sigma}(\overline{\lambda}) \leq \frac{\sigma}{(1-\overline{\lambda}\varepsilon)^3}$ (cf. Lemma 4.3), it follows that

$$\int_0^\infty e^{-t}\mathbf{P}_{\overline{\lambda}}(0 < V_n(\overline{\lambda}) \leq t)dt \quad \geq \quad M\left(\check{x}\right) - G.$$

Returning to (69), we obtain

$$\mathbf{P}(S_n > x\sigma) \quad \geq \quad 1 - \Phi\left(\check{x}\right) - G\exp\left\{-\frac{\check{x}^2}{2}\right\}.$$

Using Lemma 4.3, for all $0 \leq x \leq \sigma/(9.6\varepsilon)$, we have $0 \leq \overline{\lambda}\varepsilon \leq 1/4.8$ and

$$G \geq 26.88R\left(\overline{\lambda}\varepsilon\right)\frac{\varepsilon}{\sigma}.$$

Therefore, for all $0 \leq x \leq \sigma/(9.6\varepsilon)$,

$$\mathbf{P}(S_n > x\sigma) \quad \geq \quad 1 - \Phi\left(\check{x}\right) - 26.88R\left(\overline{\lambda}\varepsilon\right)\frac{\varepsilon}{\sigma}\exp\left\{-\frac{\check{x}^2}{2}\right\}.$$

Using the inequality $M^{-1}(t) \leq \sqrt{2\pi}\,(1+t)$ for $t \geq 0$, we get, for all $0 \leq x \leq \sigma/(9.6\varepsilon)$,

$$\mathbf{P}(S_n > x\sigma) \quad \geq \quad \Big(1 - \Phi\left(\check{x}\right)\Big)\left[1 - 67.38R\left(\overline{\lambda}\varepsilon\right)(1+\check{x})\,\frac{\varepsilon}{\sigma}\right].$$

In particular, for all $0 \leq x \leq \alpha\sigma/\varepsilon$ with $0 \leq \alpha \leq 1/9.6$, a simple calculation shows that

$$0 \leq \overline{\lambda}\varepsilon \leq \frac{2\alpha}{1+\sqrt{1-9.6\alpha}} \leq \frac{1}{4.8}$$

and

$$67.38R\left(\overline{\lambda}\varepsilon\right) \leq 67.38R\left(\frac{2\alpha}{1+\sqrt{1-9.6\alpha}}\right) \leq 67.38R\left(\frac{1}{4.8}\right) \leq 1753.23.$$

This completes the proof of Theorem 2.3.

7. Proof of Theorem 2.4

Notice that $\Psi_n'(\lambda) = T_n(\lambda) \in [0, \infty)$ is nonnegative in $\lambda \geq 0$. Let $\overline{\lambda} = \overline{\lambda}(x) \geq 0$ be the unique solution of the equation $x\sigma = \Psi_n'(\lambda)$. This definition implies that $T_n(\overline{\lambda}) = x\sigma$, $U_n(\overline{\lambda}) = \overline{\lambda} Y_n(\overline{\lambda})$ and

$$e^{-\overline{\lambda}x\sigma + \Psi_n(\overline{\lambda})} \quad = \quad \inf_{\lambda \geq 0} e^{-\lambda x\sigma + \Psi_n(\lambda)} \quad = \quad \inf_{\lambda \geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}]. \tag{70}$$

From (51), using Lemma 4.4 with $\lambda = \overline{\lambda}$ and an argument similar to (62), we obtain

$$\mathbf{P}(S_n > x\sigma) \quad = \quad \left(M\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) + \frac{26.88\theta_1\sigma^2\varepsilon}{\overline{\sigma}^3(\overline{\lambda})(1-\overline{\lambda}\varepsilon)^4} \right) \inf_{\lambda \geq 0} \mathbf{E}[e^{\lambda(S_n - x\sigma)}], \tag{71}$$

where $|\theta_1| \leq 1$. Since $M(t)$ is decreasing in $t \geq 0$ and $|M'(t)| \leq \frac{1}{\sqrt{\pi}\, t^2}$ in $t > 0$, it follows that

$$\left| M\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) - M(x) \right| \quad \leq \quad \frac{1}{\sqrt{\pi}} \frac{\left| x - \overline{\lambda}\overline{\sigma}(\overline{\lambda}) \right|}{\overline{\lambda}^2 \overline{\sigma}^2(\overline{\lambda}) \wedge x^2}. \tag{72}$$

By Lemma 4.1, we have the following two-sided bound of x:

$$\frac{(1-1.5\overline{\lambda}\varepsilon)(1-\overline{\lambda}\varepsilon)}{1-\overline{\lambda}\varepsilon + 6\overline{\lambda}^2\varepsilon^2} \overline{\lambda}\sigma \leq \frac{T_n(\overline{\lambda})}{\sigma} = x \leq \frac{1-0.5\overline{\lambda}\varepsilon}{(1-\overline{\lambda}\varepsilon)^2} \overline{\lambda}\sigma. \tag{73}$$

Using the two-sided bound in Lemma 4.3 and (73), by a simple calculation, we deduce

$$\overline{\lambda}^2 \overline{\sigma}^2(\overline{\lambda}) \wedge x^2 \geq \frac{(1-\overline{\lambda}\varepsilon)^2(1-3\overline{\lambda}\varepsilon)}{(1-\overline{\lambda}\varepsilon + 6\overline{\lambda}^2\varepsilon^2)^2} \overline{\lambda}^2 \sigma^2 \tag{74}$$

and

$$\left| x - \overline{\lambda}\overline{\sigma}(\overline{\lambda}) \right| \leq \overline{\lambda}\sigma \left(\frac{1-0.5\overline{\lambda}\varepsilon}{(1-\overline{\lambda}\varepsilon)^2} - \frac{(1-\overline{\lambda}\varepsilon)\sqrt{(1-3\overline{\lambda}\varepsilon)^+}}{1-\overline{\lambda}\varepsilon + 6\overline{\lambda}^2\varepsilon^2} \right). \tag{75}$$

From (72), (74), (75) and Lemma 4.3, we easily obtain

$$\left| M\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) - M(x) \right| \quad \leq \quad 1.11 R\left(\overline{\lambda}\varepsilon\right) \frac{\varepsilon}{\sigma}. \tag{76}$$

By Lemma 4.3, it is easy to see that

$$\frac{26.88\sigma^2\varepsilon}{\overline{\sigma}^3(\overline{\lambda})(1-\overline{\lambda}\varepsilon)^4} \leq 26.88 R\left(\overline{\lambda}\varepsilon\right) \frac{\varepsilon}{\sigma}. \tag{77}$$

Combining (76) and (77), we get, for all $0 \leq \overline{\lambda} < \frac{1}{3}\varepsilon^{-1}$,

$$M\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) + \frac{26.88\theta_1\sigma^2\varepsilon}{\overline{\sigma}^3(\overline{\lambda})(1-\overline{\lambda}\varepsilon)^4} = M(x) + 27.99\theta_2 R\left(\overline{\lambda}\varepsilon\right) \frac{\varepsilon}{\sigma}, \tag{78}$$

where $|\theta_2| \leq 1$. By (73), it follows that, for all $0 \leq \overline{\lambda} < \frac{1}{3}\varepsilon^{-1}$,

$$\overline{\lambda}\varepsilon \leq \frac{1-\overline{\lambda}\varepsilon + 6\overline{\lambda}^2\varepsilon^2}{(1-1.5\overline{\lambda}\varepsilon)(1-\overline{\lambda}\varepsilon)} x\frac{\varepsilon}{\sigma} \leq 4x\frac{\varepsilon}{\sigma}. \tag{79}$$

Implementing (78) into (71) and using (79), we obtain equality (22) of Theorem 2.4. Notice that $R < \infty$ restricts $0 \leq 4x\frac{\varepsilon}{\sigma} < \frac{1}{3}$, which implies that $0 \leq x < \frac{1}{12}\frac{\varepsilon}{\sigma}$.

[1] Arkhangelskii A N. Lower bounds for probabilities of large deviations for sums of independent random variables. Theory Probab Appl, 1989, 34: 565–575

[2] Bahadur R, Rao R R. On deviations of the sample mean. Ann Math Statist, 1960, 31: 1015–1027

[3] Bennett G. Probability inequalities for the sum of independent random variables. J Amer Statist Assoc, 1962, 57: 33–45

[4] Bentkus V. An inequality for tail probabilities of martingales with differences bounded from one side. J Theoret Probab, 2003, 16: 161–173

[5] Bentkus V. On Hoeffding's inequality. Ann Probab, 2004, 32: 1650–1673

[6] Bentkus V, Dzindzalieta D A. Tight Gaussian bound for weighted sums of Rademacher random variables. ArXiv:1307.3451, 2013

[7] Bentkus V, Kalosha N, van Zuijlen M. On domination of tail probabilities of (super)martingales: explicit bounds. Lithuanian Math J, 2006, 46: 1–43

[8] Bercu B. Inégalités exponentielles pour les martingales. Journées ALEA, 2008, 1: 1–33

[9] Bercu B, Rouault A. Sharp large deviations for the Ornstein-Uhlenbeck process. Theory Probab Appl, 2006, 46: 1–19

[10] Bernstein S N. The Theory of Probabilities. Moscow, Leningrad, 1946

[11] Bousquet O. A Bennett concentration inequality and its application to suprema of empirical processes. C R Acad Sci Paris Ser I, 2002, 334: 495–500

[12] Chaganty N R, Sethuraman J. Strong large deviation and local limit theorems. Ann Probab, 1993, 21: 1671–1690

[13] Cramér H. Sur un nouveau théorème-limite de la théorie des probabilités. Actualite's Sci Indust, 1938, 736: 5–23

[14] de la Peña V H. A general class of exponential inequalities for martingales and ratios. Ann Probab, 1999, 27: 537–564

[15] Dembo A, Zeitouni O. Large deviations techniques and applications. Springer, New York, 1998

[16] Deuschel J D, Stroock D W. Large deviations. Academic Press, Boston, 1989

[17] Eaton M L. A probability inequality for linear combination of bounded randon variables. Ann Statist, 1974, 2: 609–614

[18] Fan X, Grama I, Liu Q. About the constant in Talagrand's inequality for sums of bounded random variables. ArXiv:1206.2501, 2012, 1–22

[19] Fan X, Grama I, Liu Q. Cramér large deviation expansions for martingales under Bernstein's condition. Stochastic Process Appl, 2013, 123: 3919–3942

[20] Fan X, Grama I, Liu Q. Sharp large deviations under Bernstein's condition. C R Acad Sci Paris Ser I, 2013, 351: 845–848

[21] Fu J C, Li G, Zhao L C. On large deviation expansion of distribution of maximum likelihood estimator and its application in large sample estimation. Ann Inst Statist Math, 1993, 45: 477–498

[22] Itō K, MacKean H P. Difussion Processes and Their Sample Paths. Springer, 1996

[23] Grama I, Haeusler E. Large deviations for martingales via Cramer's method. Stochastic Process Appl, 2000, 85: 279–293

[24] Györfi L, Harremöes P, Tusnády G. Some refinements of large deviation tail probabilities. ArXiv:1205.1005v1 [math.ST], 2012

[25] Hoeffding W. Probability inequalities for sums of bounded random variables. J Amer Statist Assoc, 1963, 58: 13–30

[26] Jing B Y, Shao Q M, Wang Q. Self-normalized Cramér-type large deviations for independent random variables. Ann Probab, 2003, 31: 2167–2215

[27] Jing B Y, Liang H Y, Zhou W. Self-normalized moderate deviations for independent random variables. Sci China Math, 2012, 55(11): 2297–2315

[28] Joutard C. Sharp large deviations in nonparametric estimation. J Nonparametr Stat, 2006, 18: 293–306

[29] Joutard C. Strong large deviations for arbitrary sequences of random variables. Ann Inst Stat Math, 2013, 65(1): 49–67

[30] Rozovky L V. A lower bound of large-deviation probabilities for the sample mean under the Cramér condition. J Math Sci, 2003, 118(6)

[31] Rozovky L V. Large deviation probabilities for some classes of distributions statisfying the Cramér condition. J Math Sci, 2005, 128(1)

[32] Nagaev S V. Large deviations of sums of independent random variabels. Ann Probab, 1979, 7: 745–789

[33] Nagaev S V. Lower bounds for the probabilities of large deviations of sums of independent random variables. Theory Probab Appl, 2002, 46: 79–102; 728–735

[34] Petrov V V. Sums of Independent Random Variables. Springer-Verlag, Berlin, 1975

[35] Petrov V V. Limit Theorems of Probability Theory. Oxford University Press, Oxford, 1995

[36] Pinelis I. An asymptotically Gaussian bound on the Rademacher tails. Electron J Probab, 2012, 17(35): 1–22

[37] Pinelis I. On the Bennett-Hoeffding inequality. Ann Inst H Poincaré Probab Statist, 2014, 50(1): 15–27

[38] Rio E. A Bennett type inequality for maxima of empirical processes. Ann Inst H Poincaré Probab Statist, 2002, 6: 1053–1057

[39] Rio E. Sur la fonction de taux dans les inégalités de Talagrand pour les processus empiriques. C R Acad Sci Paris Ser I, 2012, 350: 303–305

[40] Rio E. Inégalités exponentielles et inégalités de concentration. Institut Mathématique de Bordeaux, 2012, 1: 1–22.

[41] Saulis L, Statulevicius V A. Limit theorems for large deviations. Springer, 1991

[42] Shevtsova I G. An improvement of convergence rate estimates in the Lyapunov theorem. Doklady Math, 2010, 82: 862–864

[43] Sakhanenko A I. Berry-Esseen type bounds for large deviation probabilities. Siberian Math J, 1991, 32: 647–656

[44] Shao Q M. A Cramér type large deviation result for Student's t-statistic. J Theoret Probab, 1999, 12(2): 385–398

[45] Talagrand M. The missing factor in Hoeffding's inequalities. Ann Inst H Poincaré Probab Statist, 1995, 31: 689–702

[46] van de Geer S. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting process. Ann Statist, 1995, 23: 1779–1801

[47] van de Geer S, Lederer J. The Bernstein-Orlicz norm and deviation inequalities. Probab Theory Relat Fields, 2013, 157: 225–250